

## **REMARKS**

This paper is submitted in response to the Office Action issued by the United States Patent and Trademark Office (USPTO) on May 7, 2009 with respect to the application under consideration. Examiner Mr. Adrian L. Kennedy is thanked for the thorough examination and search of the subject application. Applicants respectfully submit a response thereto.

Applicants amend claims 1-5, 7-17, 19-31, and 33-38, cancel claims 6, 18 and 32, and add new claims 39-42. Claims 1-5, 7-17, 19-31, and 33-38 are amended to more distinctly and clearly define the invention therein. Support for the amendments and new claims can be found in the specification and drawings of the present application and Applicants respectfully submit that no new matter is added.

Applicants note that in the May 7, 2009 Office Action, Examiner has rejected pending claims 1 to 38 under 35 U.S.C § 101. Pending claims 1-6, 8-10, 13-18, 20-22, 25-32 and 34-36 are also rejected under 35 U.S.C § 103(a) as being unpatentable over He et al. (hereinafter He) (Machine Learning Methods for Chinese Web Page categorization) in view of US Patent No. 5,297,039 to Kanaegami et al. (hereinafter Kanaegami).

Applicants also note that Examiner has rejected pending claims 7, 19 and 33 under 35 U.S.C § 103(a) as being unpatentable over He in view of Kanaegami and further in view of Tan et al., Predictive Self-Organizing Networks for Text Categorization (hereinafter "Tan 1").

Applicants further note that Examiner has rejected pending claims 11, 12, 23, 24, 37 and 38 under 35 U.S.C § 103(a) as being unpatentable over He in view of Kanaegami and further in view of Tan et al., Learning User Profiles for Personalized Information Dissemination (hereinafter "Tan 2").

### **Rejections under 35 U.S.C § 101 for claims 1 to 38**

In the Office Action, Examiner rejects pending claims 1 to 38 as being unpatentable under 35 U.S.C §101. In response, Applicant has amended claim 1 as follows:

A method for discovering knowledge from a set of text documents using a processor, the method comprising:

extracting semi-structured meta-data from the set of text documents using a meta-data extractor, the semi-structured meta-data comprising a plurality of concepts and a plurality of relations between the concepts;

filtering the semi-structured meta-data to identify a set of key concepts and a corresponding set of key relations between the key concepts, the set of key concepts corresponding to the plurality of concepts;

deriving at least one set of sub-concepts corresponding to the set of key concepts based upon data within a domain knowledge base, using a meta-data transformer;

formulating a plurality of training samples, each training sample including a vector representing a sub-concept and a vector representing a key concept; and

analyzing the plurality of training samples using an associative discoverer to derive a set of associations between a set of vectors representing a sub-concept and at least one vector representing a key concept,

wherein neither the set of text documents nor the semi-structured meta-data mention the set of associations, and

wherein the set of associations corresponds to discovered knowledge that is extractable by a knowledge interpreter.

Thus, the method of amended claim 1 recites (emphasis added) the discovery of knowledge from a set of text documents using a processor. The method comprises extracting semi-structured meta-data from the set of text documents using a meta-data extractor, filtering the semi-structured meta-data to identify a set of key concepts and a corresponding set of key relations between the key concepts and deriving at least one set of sub-concepts corresponding to the set of key concepts based upon data within a domain knowledge base using a meta-data transformer.

The method also comprises formulating a plurality of training samples and analyzing the plurality of training samples using an associative discoverer to derive a set of associations between a set of vectors representing a sub-concept and at least one vector representing a key concept. Neither the set of text documents nor the semi-structured meta-data mention the set of

associations, and the set of associations corresponds to discovered knowledge that is extractable by a knowledge interpreter.

Applicant respectfully submit that amended claim 1 is directed to statutory subject matter and is patentable under 35 U.S.C § 101.

In Ex parte Bilski, Federal Circuit held that, to be statutory, a process must either:

1. be tied to a particular machine; or
2. transform a particular article into a different state or thing.

Applicant submits that claim 1 as amended meets both of the tests set forth for statutory subject matter in Ex parte Bilski.

Addressing the first part of the two-pronged test (i.e., tied to a particular machine), Applicant submits that amended claim 1, as presented above, recites (emphasis added):

- The discovery of knowledge from a set of text documents using a processor
- Extraction of semi-structured meta-data from the set of text documents using a meta-data extractor
- Derivation of at least one set of sub-concepts using a meta-data transformer
- Analyzing the plurality of training samples using an associative discoverer
- Discovered knowledge that is extractable by a knowledge interpreter

Therefore, Applicants submit that amended claim 1, wherein a processor is used to discover knowledge from a set of documents, is indeed tied to a particular machine. Furthermore, amended claim 1 further recites tangible processing components (i.e., the meta-data extractor, the meta-data transformer, and the associative discoverer) which put the processing steps recited in amended claim 1 into effect. Additionally, it is further recited in amended claim 1 that the discovered knowledge is extractable by a knowledge interpreter.

Moreover, pages 19 and 20 of Applicant's specification (paragraphs 87 – 90 in corresponding published Patent Application No. 20060026203) clearly and fully tie each of the processor, the meta-data extractor, the meta-data transformer, the associative discoverer, and the knowledge interpreter recited in amended claim 1 to a particular machine.

Applicants further submit that amended claim 1 is directed to a method that transform a particular article into a different state or thing. Claim 1 recites:

- extracting semi-structured meta-data from the set of text documents
- filtering the semi-structured meta-data to identify a set of key concepts and a corresponding set of key relations between the key concepts
- deriving at least one set of sub-concepts corresponding to the set of key concepts based upon data within a domain knowledge base
- formulating a plurality of training samples
- analyzing the plurality of training samples to derive a set of associations between a set of vectors representing a sub-concept and at least one vector representing a key concept.
- the set of associations corresponds to discovered knowledge that is extractable by a knowledge interpreter

Applicants submit that in amended claim 1, 'raw data' (i.e., semi-structured meta-data) is first extracted from the set of text documents. The 'raw data' is refined (i.e., filtering the semi-structured meta-data) and a set of sub-concepts is derived from the refined 'raw data' (i.e., deriving at least one set of sub-concepts corresponding to the set of key concepts based upon data within a domain knowledge base). A plurality of training samples is then formulated from the refined 'raw data' and the derived set of sub-concepts (i.e., formulating a plurality of training samples). A set of associations, corresponding to discovered knowledge, is derived by analyzing the plurality of training samples.

Therefore, it is firstly recited in amended claim 1 that 'raw data' (i.e., semi-structured meta-data extracted from the set of text documents), after refining via filtering (i.e., filtering the semi-structured meta-data), is transformed into a set of sub-concepts. It is secondly recited in amended claim 1 that the set of sub-concepts is further transformed into a plurality of training samples (i.e., formulating a plurality of training samples). The plurality of training samples are analyzed to derive a set of associations that is not mentioned in the set of text documents nor the semi-structured metadata, and which corresponds to discovered knowledge. Amended claim 1 clearly recites a transformative process.

From the foregoing, Applicant respectfully reiterates that amended claim 1 fulfils both of the tests for statutory subject matter set forth in *Ex parte Bilski*. Amended claim 1 is hence statutory subject matter and is patentable under 35 U.S.C § 101.

The above remarks made in relation to amended claim 1 apply analogously to amended claims 13 and 25.

Specifically, the claimed invention as recited in amended claims 13 and 25, recite tangible and concrete processing steps to obtain tangible and concrete results. As above, the semi-structured meta-data are transformed to sub-concepts, from which training samples are formulated and analyzed to obtain a set of associations between a set of vectors representing a sub-concept and at least one vector representing a key concept. The set of associations represents a transformation of the raw data into new or discovered information, corresponding to discovered knowledge which is extractable by a knowledge interpreter.

Hence Applicant submits that there is a practical application of tangible and concrete processing steps to transform the semi-structured meta-data to sub-concepts; formulate by way of further transformation the training samples; and transformatively analyze the training samples to derive the set of associations between a set of vectors representing a sub-concept and at least one vector representing a key concept. As disclosed in page 8 lines 9 to 23 of the specification (paragraphs 48 and 49 of corresponding published Patent Application No. 20060026203), the set of associations is tangible and has real world application(s).

Pending claims dependent upon respective independent pending claims 1, 13 and 25 have been amended for consistency and are submitted to be patentable under 35 U.S.C § 101.

From the foregoing remarks, Applicant respectfully reiterates that the claimed invention is statutory subject matter and respectfully requests withdrawal of Examiner's rejections under 35 U.S.C § 101.

**Obviousness rejections under 35 U.S.C § 103(a) for claims 1-6, 8-10, 13-18, 20-22, 25-32 and 34-36**

In the Office Action, Examiner rejects pending claims 1-6, 8-10, 13-18, 20-22, 25-32 and 34-36 under 35 U.S.C § 103(a) as being unpatentable over He in view of Kanaegami.

Claim 1 as amended recites the following:

A method for discovering knowledge from a set of text documents using a processor, the method comprising:

extracting semi-structured meta-data from the set of text documents using a meta-data extractor, the semi-structured meta-data comprising a plurality of concepts and a plurality of relations between the concepts;

filtering the semi-structured meta-data to identify a set of key concepts and a corresponding set of key relations between the key concepts, the set of key concepts corresponding to the plurality of concepts;

deriving at least one set of sub-concepts corresponding to the set of key concepts based upon data within a domain knowledge base, using a meta-data transformer;

formulating a plurality of training samples, each training sample including a vector representing a sub-concept and a vector representing a key concept; and

analyzing the plurality of training samples using an associative discoverer to derive a set of associations between a set of vectors representing a sub-concept and at least one vector representing a key concept,

wherein neither the set of text documents nor the semi-structured meta-data mention the set of associations, and

wherein the set of associations corresponds to discovered knowledge that is extractable by a knowledge interpreter.

Thus, the method of amended claim 1 recites extracting semi-structured meta-data from the set of text documents using a meta-data extractor, the semi-structured meta-data comprising a plurality of concepts and a plurality of relations between the concepts; filtering the semi-structured meta-data to identify a set of key concepts and a corresponding set of key relations between the key concepts, the set of key concepts corresponding to the plurality of concepts and deriving at least one set of sub-concepts corresponding to the set of key concepts based upon data within a domain knowledge base, using a meta-data transformer.

Pending claim 1 further recites formulating a plurality of training samples, each training sample including a vector representing a sub-concept and a vector representing a key concept; and analyzing the plurality of training samples using an associative discoverer to derive a set of associations between a set of vectors representing a sub-concept and at least one vector representing a key concept. Neither the set of text documents nor the semi-structured meta-data

mention the set of associations. The set of associations corresponds to discovered knowledge that is extractable by a knowledge interpreter.

Applicants respectfully submit that amended claim 1 is patentably distinct over He in view of Kanaegami.

Examiner asserts that the difference between the claimed invention and He is the teaching of “pairs of key entities” and a “plurality of attributes attributed thereto” in the claimed invention. Examiner further asserts that such a feature is disclosed in Kanaegami.

Regardless of Examiner’s assertion of any apparent equivalence between particular elements of Applicant’s pending claim 1 and He and/or Kanaegami, Applicants submit that key differences exist between pending claim 1 and any teaching or suggestion of He and/or Kanaegami.

In particular, with regard to He, Applicants submit that He discloses a method for text categorization for Chinese information by the application of three known statistical machine-learning methods to Chinese web page categorization. The three statistical machine-learning methods are namely k Nearest Neighbor system (kNN), Support Vector Machines (SVM) and Adaptive Resonance Associative Map (ARAM). He investigates the capabilities of these methods in learning categorization knowledge from real-life web documents. In addition, He further investigates whether the incorporation of domain knowledge derived from the category description can enhance ARAM’s predictive performance.

Therefore, He relates to text categorization in which one or multiple predefined category labels are automatically assigned to free text documents (pg 93 lines 1 to 5).

In contrast, amended claim 1 recites:

A method for discovering knowledge from a set of text documents using a processor, the method comprising:

...

deriving at least one set of sub-concepts corresponding to the set of key concepts based upon data within a domain knowledge base, using a meta-data transformer;

formulating a plurality of training samples, each training sample including a vector representing a sub-concept and a vector representing a key concept; and  
analyzing the plurality of training samples using an associative discoverer to derive a set of associations between a set of vectors representing a sub-concept and at least one vector representing a key concept,

wherein neither the set of text documents nor the semi-structured meta-data mention the set of associations, and

wherein the set of associations corresponds to discovered knowledge that is extractable by a knowledge interpreter.

Applicants respectfully disagree with Examiner's asserted equivalences between the claimed invention and He, as He relates to text categorization in which one or multiple predefined category labels are automatically assigned to free text documents whereas the claimed invention relates to the discovery of knowledge from a set of text documents.

Furthermore, although He discloses learning of categorization knowledge from real-life web documents and the incorporation of domain knowledge to enhance an ARAM's predictive performance (pg 93 lines 20 to 30), He fails to disclose the claimed invention's derivation of sub-concepts corresponding to the key concepts using a domain knowledge base for the purpose of formulating training samples that involve sub-concepts and key concepts.

With respect to such training samples, amended claim 1 recites:

formulating a plurality of training samples, each training sample including a vector representing a sub-concept and a vector representing a key concept; and  
analyzing the plurality of training samples using an associative discoverer to derive a set of associations between a set of vectors representing a sub-concept and at least one vector representing a key concept  
wherein neither the set of text documents nor the semi-structured meta-data mention the set of associations, and



wherein the set of associations corresponds to discovered knowledge that is extractable by a knowledge interpreter

Applicants submit that the training samples are used, in the claimed invention, for input to the associative discoverer to derive or generate associations corresponding to discovered knowledge. In contrast, the ARAM in He generates recognition categories from the input training patterns (pg 99 lines 38 to 40) by simply accepting (at  $F_1^a$  and  $F_1^b$ ) a document feature vector and a class prediction vector, as input training patterns, and associating each category with its respective prediction to formulate recognition categories of input patterns (pg 96 lines 1 to 10).

Furthermore, He fails to disclose any type of analysis of such training samples using an associative discoverer to derive a set of associations between a set of vectors representing a sub-concept and at least one vector representing a key concept, particularly any set of associations that the original source documents and semi-structured meta-data fail to mention. Hence, Applicants respectfully submit that nowhere does He recognize or contemplate that something can be done to process any type of input training patterns to discover knowledge in the manner recited in the claimed invention.

Applicant respectfully reiterates that amended claim 1 requires “deriving at least one set of sub-concepts corresponding to the set of key concepts based upon data within a domain knowledge base, using a meta-data transformer.” Therefore, the domain knowledge base as recited in amended claim 1 facilitates the derivation of at least one set of sub-concepts which is subsequently used for the formulation of training samples. In contrast, the domain knowledge in He simply generates additional rules for each category (pg 98 lines 20 to 25) for the purpose of improving the ARAM’s performance (pg 100 lines 8 to 11), and does not relate to the processing of input training patterns to derive sub-concepts in conjunction with key concepts, as recited in the claimed invention.

He is directed to the improvement of ARAM performance. Specifically, He discloses (at pg 100 lines 8 to 11) that an improvement in performance is more likely to be observed where:

- 1) categories are well defined; and

- 2) there are relatively fewer numbers of training patterns.

Therefore, while He may inform a person of ordinary skill about a manner of using a knowledge database to attempt to improve the performance of an ARAM, He fails to provide any guidance or incentive to any person of ordinary skill in the art with regard to deriving sub-concepts, formulating a plurality of training samples, or analyzing the plurality of training samples to generate associations corresponding to discovered knowledge that is not mentioned in the set of text documents or the semi-structured meta-data, as recited by amended claim 1.

With regard to Kanaegami, Applicant respectfully disagrees with Examiner's assertions in respect of the "elements of the triplet".

In particular, the "elements of the triplet" taught by Kanaegami are simply elements that are readily perceivable in the analysis network (column 13 lines 1 to 10). In contrast, the set of sub-concepts as recited in amended claim 1 is derived from the key concepts, based upon data within a domain knowledge base, and used thereafter to formulate training samples.

As such, Kanaegami's "elements" are already or *a priori* perceived in the analysis network and are not equivalent to, nor suggestive of, nor an incentive for recognizing any need to derive or manner of deriving sub-concepts, training samples, or associations as recited in amended claim 1.

Accordingly, He and Kanaegami, individually or in combination, fail to result in the claimed invention. Furthermore, the He and Kanaegami, individually or in combination, fail to even partially guide or lead a person of ordinary skill in the art in any manner toward a method for generating discovered knowledge involving the derivation of associations that are not mentioned in either of input text documents or semi-structured meta-data extracted from the text documents in the manner recited by amended claim 1.

Remarks made above in relation to amended claim 1 analogously apply to amended claims 13 and 25.

Accordingly, with regard to Examiner's rejections under 35 U.S.C. § 103(a) of pending claims 1, 13 and 25, Applicants submit that these rejections are consequently disposed of and pending claims 1, 13 and 25 as amended are in condition for allowance. Applicants submit that other 35 U.S.C. § 103(a) rejections for pending dependent claims 2 to 5, 8 to 10, 14 to 17, 20 to

22, 26 to 31 and 34 to 36 are consequently disposed of and therefore such pending dependent claims are in condition for allowance. Applicants respectfully request withdrawal of Examiner's rejections under 35 U.S.C. § 103(a).

**Rejection of Claims 7, 19 and 33 under 35 U.S.C. § 103(a)**

In the Office Action, Examiner has rejected dependent pending claims 7, 19 and 33 under 35 U.S.C § 103(a) as being unpatentable over He in view of Kanaegami and further in view of Tan 1.

Applicants respectfully submit that the invention for discovering knowledge from text documents as recited in amended independent claims 1, 13 and 25 is not disclosed by any portion of He, Kanaegami, and/or Tan 1. Applicants note that the inclusion of Tan 1 with He and/or Kanaegami fails to give rise to a method for generating discovered knowledge involving the derivation of associations that are not mentioned in either of input text documents or semi-structured meta-data extracted from the text documents in the manner recited by amended claim 1.

No combination of He, Kanaegami, and Tan 1 results in or leads to the invention of amended claims 1, 13, and 25. No combination of He, Kanaegami, and Tan 1 provides any insight into any manner of or incentive for deriving or attempting to derive any associations that are not mentioned in either of input text documents or semi-structured meta-data as recited in claim 1. Hence, the invention as recited in amended claims 1, 13 and 25 would not have been available or obvious to a person skilled in the art at the time of the claimed invention's filing date.

Applicant respectfully submits that Examiner's rejections of the respective dependent claims in relation to Tan are rendered moot.

**Rejection of Claims 11, 12, 23, 24, 37 and 38 under 35 U.S.C. § 103(a)**

In the Office Action, Examiner has rejected pending dependent claims 11, 12, 23, 24, 37 and 38 under 35 U.S.C § 103(a) as being unpatentable over He in view of Kanaegami and further in view of Tan 2.

Applicant respectfully submits that the invention for discovering knowledge from text documents as recited in amended independent claims 1, 13 and 25 is not disclosed by any portion

of He, Kanaegami, and/or Tan 2. Application submits that no combination of He, Kanaegami, and/or Tan 2 yields a method for generating discovered knowledge involving the derivation of associations that are not mentioned in either of input text documents or semi-structured meta-data extracted from the text documents in the manner recited by amended claim 1.

No portion of He, Kanaegami, and Tan 2 considered alone or in combination results in or leads to the invention of amended claims 1, 13, and 25. He, Kanaegami, and Tan 2 fail to provide any insight into any manner of or incentive for deriving or attempting to derive any associations that are not mentioned in either of input text documents or semi-structured meta-data as recited in claim 1. Hence, the invention as recited in amended claims 1, 13 and 25 would not have been available or obvious to a person skilled in the art at the time of the claimed invention's filing date.

Hence Applicant respectfully submits that Examiner's assertions of the respective dependent claims in relation to Tan are rendered moot.

#### **New Claims**

Applicants have added new claims 39 - 42, which are fully supported by the originally filed specification and figures. Applicants submit that new claims 39 - 42 are novel as well as nonobvious over any combination of He, Kanaegami, Tan 1, and Tan 2.

#### **Conclusion**

In accordance with the foregoing remarks, Applicants request withdrawal of rejections of pending claims 1 to 38 under 35 U.S.C. § 103(a). Examiner reconsideration and issuance of a Notice of Allowance are hereby respectfully requested. In the event that an extension of time is necessary to allow for consideration of this paper, such extensions are hereby petitioned. The Office is also authorized to charge deposit account for any fees owed.

Respectfully submitted,

/Jonathan M. Harris/

---

Jonathan M. Harris  
PTO Reg. No. 44,144  
Conley Rose, P.C.  
(713) 238-8000 (Phone)  
(713) 238-8008 (Fax)  
ATTORNEY FOR APPLICANTS